1 Warmups

1.1 Sum of Random Variables

1. What is the distribution (with name and parameter(s)) of the average of \( n \) i.i.d. random variables, \( X_1, \ldots, X_n \), each with mean \( \mu \) and variance \( \sigma^2 \)?

2. For each \( X \) and \( Y \) below, let \( X \) and \( Y \) be independent.
   
   a. Let \( X \sim N(\mu_1, \sigma_1^2) \) and \( Y \sim N(\mu_2, \sigma_2^2) \). What is \( \mu \) and \( \sigma^2 \) for \( X + Y \sim N(\mu, \sigma^2) \)?
   
   b. Let \( X \sim \text{Uni}(0, 1) \) and \( Y \sim \text{Uni}(0, 1) \). What is the PDF for \( X + Y \)?
   
   c. In general, two random variables \( X \) and \( Y \), what is the PDF \( f \) of \( X + Y \)?

1. According to the central limit theorem, this can be modeled as \( N(\mu, \sigma^2/n) \).
   
   a. \( \mu = \mu_1 + \mu_2 \) and \( \sigma^2 = \sigma_1^2 + \sigma_2^2 \). How convenient!
   
   b. \( f_{X+Y}(a) = \begin{cases} 
   a & 0 \leq a \leq 1 \\
   2 - a & 1 \leq a \leq 2 \\
   0 & \text{otherwise} \end{cases} \)
   
   c. \( f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy \)

   It is good to remember these equations, but perhaps another message from lecture that it is difficult to sum random variables. The derivation for Uniform distributions is difficult. And solving for the general random variables is even worse. But we can pick distributions, like the Normal distribution, that are easy to use!

1.2 Sample and Population Mean

Computing the sample mean is similar to the population mean: sum all available points and divide by the number of points. However, sample variance is slightly different from population variance.

1. What is the difference between the population variance, \( \sigma^2 \), and sample variance, \( S^2 \)? What is the difference between the sample variance, \( S^2 \), and variance of the sample mean, \( \text{Var}(\bar{X}) \)?
2. Consider the equation for population variance, and an analogous equation for sample variance.

\[ \sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2 \]  

(1)

\[ S_{biased}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \]  

(2)

\[ S_{biased}^2 \] is a random variable to estimate the constant \( \sigma^2 \). Because it is biased, \( E[S_{biased}^2] \neq \sigma^2 \). Is \( E[S_{biased}^2] \) greater or less than \( \sigma^2 \)?

3. Consider an alternative Random Variable, \( S_{unbiased}^2 \) (known simply as \( S^2 \) in class). The technique of unbiasing variance is known as Bessel’s correction. Write the \( S_{unbiased}^2 \) equation.

4. (Optional) Can you think of any examples in which the biased estimator might be preferred over the unbiased estimator?

1. Population variance, \( \sigma^2 \), is the true variance of a population (or random variable). Sample variance, \( S^2 \), is the unbiased estimate of true variance based on a random subsample. Variance of sample mean, \( \text{Var}(\bar{X}) \), is the amount of spread in the estimation of the true mean.

2. \( E[S_{biased}^2] < \sigma^2 \). The intuition is that the spread of a sample of points is generally smaller than the spread of all the points considered together. This becomes more clear when we consider the unbiased version and how it makes the expression evaluate to a larger number.

3. \[ S_{unbiased}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \]

4. Some might prefer the biased estimator over the unbiased estimator when it helps them skew statistics favorably for them. For example, a team in a legal case may use the biased estimator over the unbiased if the bias is in their favor.

In addition, often times in machine learning, we are concerned with minimizing mean squared error (which is calculated using a combination of the variance and bias). Sometimes, by slightly increasing bias, we are able to greatly decrease variance, and therefore minimize our mean squared error in our ML models. Thus, we may prefer a biased estimator in this case. This is outside the scope of our class, but it’s interesting to think about!

### 1.3 Beta

1. Suppose you have a coin where you have no prior belief on its true probability of heads \( p \). How can you model this belief as a Beta distribution?

2. Suppose you have a coin which you believe is fair, with “strength” \( \alpha \). That is, pretend you’ve seen \( \alpha \) heads and \( \alpha \) tails. How can you model this belief as a Beta distribution?
3. Now suppose you take the coin from the previous part and flip it 10 times. You see 8 heads and 2 tails. How can you model your posterior belief of the coin’s probability of heads?

1. Beta(1, 1) is a uniform prior, meaning that prior to seeing the experiment, all probabilities of heads are equally likely.
2. Beta(\(\alpha + 1\), \(\alpha + 1\)). This is our prior belief about the distribution.
3. Beta(\(\alpha + 9\), \(\alpha + 3\))

2 Problems

2.1 Beta Sum

What is the distribution of the sum of 100 i.i.d. Betas? Let \(X\) be the sum

\[
X = \sum_{i=1}^{100} X_i \quad \text{where each } X_i \sim \text{Beta}(a = 3, b = 4)
\]

Note the variance of a Beta:

\[
\text{Var}(X_i) = \frac{ab}{(a+b)^2(a+b+1)} \quad \text{where } X_i \sim \text{Beta}(a, b)
\]

By the Central Limit Theorem, the sum of equally weighted IID random variables will be Normally distributed. We calculate the expectation and variance of \(X_i\) using the beta formulas:

\[
E(X_i) = \frac{a}{a+b} = \frac{3}{7} \approx 0.43
\]

\[
\text{Var}(X_i) = \frac{ab}{(a+b)^2(a+b+1)} \quad \text{Variance of a Beta}
\]

\[
= \frac{(3+4)^2(3+4+1)}{3 \cdot 4} = \frac{12}{49 \cdot 8} \approx 0.03
\]

\[
X \sim N(\mu = n \cdot E[X_i], \sigma^2 = n \cdot \text{Var}(X_i)) \sim N(\mu = 43, \sigma^2 = 3)
\]
2.2 Food for Thought
Karel the dog eats an unpredictable amount of food. Every day, the dog is equally likely to eat between a continuous amount in the range 100 to 300 grams. How much Karel eats is independent of all other days. You only have 6.5kg of food for the next 30 days. What is the probability that 6.5kg will be enough for the next 30 days?

The distribution of the sum is given by the central limit theorem. Let \( X_i \sim Uni(100, 300) \) where \( E[X_i] = 200 \) and \( Var(X_i) = \frac{1}{12}(200)^2 \approx 3333 \).

\[
Y = \sum_i X_i
\]

Let’s approximate \( Y \) with a normal random variable.

\[
Y \sim N(6000, 316.212^2)
\]

Let’s now calculate our final probability

\[
P(Y < 6500) = F_Y(6500)
= P \left( Z < \frac{6500 - 6000}{316.212} \right)
= P(Z < 1.58)
= 0.943
\]